Beutnagel 3-12-9

REMARKS

A number of typographical errors are corrected in the specification.

Claims 23-34 were rejected under 35 USC 112, first paragraph because, according to the Examiner, the claims contain subject matter that was not described in the specification. In particular, the Examiner asserts that the phrases "parameter information tuples N," "N tuples," and "parameter tuples," are "nowhere found in the written description of the specification, and therefore is not reasonably conveyed to one skilled in the art at the time the application was files had possession of the claimed invention.

Applicants respectfully traverse.

Clearly, the word that the Examiner has a problem with is "tuple," and indeed, this word is "nowhere found in the written description." However, the word tuple is a common word in the arts that involve storage or handling of information. A quick reference to web site http://whatis.techtarget.com[1], which contains technical definitions reveals that the term "tuple" means to "an ordered set of values." Often, fields that contain the values are allowed to have any length, and in such applications a chosen delimiter separates the fields. Tuples can contain a mixture of data types; for example, one field may contain numbers, another field may contain text, a third field may contain dates, etc. It is noted that since the term "tuple" is used most often in the applications where data is stored, it is implied that a tuple has a finite, though no necessarily fixed, number of fields.

In the case at hand, the collection of fields that describe a TTS sentence is a tuple. Illustratively, the tuple can contain only 4 fields, or values (as described at page 1, line 30 to page 2, and the first 4 lines of page 3):

the TTS_Sentence_Start_Code, (32 bits),

the TTS_Sentence_ID value (10 bits),

a Silence flag (1 bit), and

a Silence_duration value (12 bits).

The tuple can also contain more fields (when the Silence flag indicates a non-silence interval).

---

[1] More particularly, URL http://whatis.techtarget.com/definition/0,,sid9_gci213231,00.html

Beutnagel 3-12-9

It is noted that the tuple described in the specification is one where the fields have fixed lengths and, therefore, they do not need explicit delimiters (thereby saving bits).

As explained in the definition of the term "tuple" in the above-identified Internet site of technical definitions, a tuple can be nested within a tuple. Hence, it is not strange to characterize the information described at page 5 of the specification (lines 6-27) as a tuple. This information can be viewed to contain a first part and a second part. The first part contains:

- A 1 bit field for the Dur_Enable value;
- A 1 bit field for the F0_Contour_Enable value;
- A 1 bit field for the Energy_Contour_Enable value;
- A 10 bit field for the Number_of_Phonemes value; and
- A 13 bit field for the Phonemes_Symbols_Length value.

A review of claim 23 shows that this first part clearly corresponds to the "preface information that includes indication of number of phonemes" limitation of claim 1.

The second part contains the fields specifies at page 5, lines 14 – 26, and within this information there is a 5 bits num_F0 field that is followed by a number of 2-field strings, or tuples. Those are the F0_Contour_Each_Phoneme 8-bit field, and the F0_Each_Phoneme_time 12-bit field.

Applicants respectfully submit that person who has reasonable skill in the art to which the instant invention pertains would understand the correspondence between the two-field string comprising the F0_Contour_Each_Phoneme 8-bit field followed by the F0_Each_Phoneme_time 12-bit field as a "tuple," since it clearly represents an "ordered set of values." Put differently, applicants respectfully submit that it would have been perfectly permissible to define the two concatenated fields as a string of ordered values and, therefore, it is equally permissible to define the fields by the equivalent term, i.e., "tuple."

Of course, it follows that if it is permissible to refer to the strings of the F0_Contour_Each_Phoneme field followed by the F0_Each_Phoneme_time field as "tuples," it is also permissible to refer to the num_F0 field as the "indication of number of parameter information tuples, N."

Beutnagel 3-12-9

It is respectfully submitted that the above remarks traversing the rejection of claim 23 based on 35 USC 112, first paragraph apply with equal vigor to claims 23-34.

Claims 1-5, 7, 10, 13-22 were rejected under 35 USC 103 as being unpatentable over Yang et al, US Patent 5,970,459 in view of Campbell et al, US Patent 6,366,883. Applicants respectfully traverse.

Actually, the Examiner's explanation of the rejection can be divided into two groups. The first group includes claims 1 and 18-22, and the second group includes claims 2-5, 7, 10, 13, and 14.

The remarks in the first group make reference to a Lee et al reference. It is assumed that the rejection of claims 1 and 18-22 was probably intended to be over Yang et al in view of Campbell et al **and** Lee et al, and it is so assumed. Since the identity of this reference is not provided, it is further assumed that it is the same Lee et al reference that was employed by the Examiner in previous Office Actions. Formally, applicants respectfully traverse the rejection of claims 1 and 18-22 under 35 USC 103 over Yang et al in view of Campbell et al **and** Lee et al.

With reference to claim 1, it is noted that Yang et al teach that a language processing unit 1 converts an input text to a phoneme string. This unit also estimates prosodic information and "symbolizes it." It is not clear what Yang at el mean by "symbolizes it." Probably a collection of symbols is developed; but this is only a surmise. According to col. 1, lines 37-40, "[T]he symbol of the prosodic information is estimated from the phrase boundary, clause boundary, accent position, sentence patterns, etc." Apparently, Yang et al expect the input text to comprise sentences that include phrases and clauses, with accent position and sentence patterns that somehow can be determined. How this is done is not described. Continuing the perusal of the col. 1 text of Yang et al, it is understood that a prosody processing unit 2 "calculates the values for prosody control parameters from symbolized prosodic information by using rules and tables." As best understood, this means that unit 1 attaches a symbol to prosody information derived from the text, and unit 2 creates "prosody control parameters" from the prosody symbols, using rules and tables. Continuing, according to col. 1, lines 43-45 "[T]he prosody control parameters includes phone duration and pause interval information," while according to col. 4, lines 63-66, [T]he prosodic control parameter

includes the time duration of phonemes, contour of pitch, contour of energy, position of pause, and length." It is not clear what the "pause" relates to, but it is surmised to be a pause following a phoneme. The "length" probably refers to the length of the pause, but it is not clear what is meant by the "position of pause." Applicants respectfully reiterate that the Yang et al disclosure is not clear relative to a number of notions mentioned and, therefore, the teaching is faulty, and cannot be used even if by some interpretation and extension it were to teach applicant's invention to the extent asserted by the Examiner.

The Examiner asserts that "a third step including at least one of said phonemes at a time offset from the beginning of the duration of said phoneme that is greater than zero less than the duration of said phoneme" is taught by Yang et al at col. 5 lines 1-12 as "offset adjustment of the time duration of the phonemes." Applicants respectfully disagree substantively, but for now, with reference to claim 1, applicants simply note that claim 1 does not include the notion of a time offset at all.

The Examiner admits that Lee et al do not explicitly teach "any selected point in time for reaching said target value." However, the Examiner asserts that Campbell et al do teach "a selected point in time for reaching a target value, citing col. 16, line 14 through col. 17, line 23. Applicants respectfully disagree.

The Campbell arrangement comprises an analyzer and a synthesizer. The analysis is performed once, when data is first presented, and the synthesis is performed many times (potentially) – each time an output of speech is desired.

FIG. 4 of Campbell is a flowchart of the speech analysis process (see col. 15, line 54), and a first portion of the passage cited by the Examiner (col. 16, lines 14 – 33) addresses steps in the analysis process. For this reason alone, the passage relating to FIG. 4 is immaterial to claim 1, since claim 1 defines a method for generating a signal rich in prosody information (i.e., a speech synthesis method).

Aside from this reason, it is noted that relative to phonemes, all that this passage teaches is that the analysis process determines the start time and the duration of each phoneme, and this information is stored in a file. It certainly does not specify a target value, and a selected point in time of reaching the target value, as asserted by the Examiner. The only possibility that comes to mind is that the Examiner considers the instant that corresponds to the start time, plus the duration (i.e., the end of the phoneme)

to be a specification of "a selected point in time of reaching the target value." If that is what the Examiner has in mind, then applicants respectfully traverse because claim 1 already specifies the duration of a phoneme (in the second clause), and the specification of "any selected point in time" in claim 1 is a specification over and above the duration of the phoneme. The references simply do not have or suggest a specification of a phoneme duration and another specification of some other time for a prosodic feature of a phoneme.

Although applicants believe that claim 1 in its unamended form is clearly not suggested by the combination of Yang et al, Lee et al, and Campbell et al, in the interest of a clearer expression of the claimed subject matter claim 1 is amended to specify that the point in time for reaching the target value of a prosody parameter of a phoneme is a "point in time that is unrestricted to any particular point within said duration." That is, the claimed method, as amended, is clearly one that allows the specification of the reaching-the-target value of a prosody parameter to be anywhere within the duration of a phoneme. Applicants believe that this amendment does not change the claim relative to the clarity by which the claim overcomes the prior art. It is only a clearer expression of the invention in general.

The remainder portion of the passage cited by the Examiner (col. 16, line 34 to col. 17, line 23) is directed to FIGS. 5 and 6, which are flowcharts of the weighting coefficient training process. The weighting coefficients are the 16-degree cepstrum coefficients that are mentioned in col. 2 line 22-24, and although they are time-normalized using prosodic rules, it does not mean that they are prosodic parameters. Moreover, even if one were to assume that they are, it is still true that none of the steps described in connection with FIGS. 5 and 6 describes or suggests any specification of a time, or duration, of a prosodic feature of a phoneme; and certainly nothing is found in the method described in connection with FIGS. 5 and 6 that describes or suggests any specification of a time, or duration, of a prosodic feature of a phoneme separate from, or in addition to, the specification of a phoneme's duration. There is no notion of a target value, there is no concept of a time for reaching the target value, and there is no notion of specifying a time for reaching the target value. This is true in connection with an

"feature parameter" and, equally true in connection with prosody parameters, since they are not mentioned at all in the first place.

If the Examiner believes otherwise, applicants respectfully request that the Examiner point to any one, or a collection of steps, in FIGS. 5 and 6 (or anywhere else) that specifies a step of:

> inserting, for at least one of said phonemes, at least two prosody parameter specifications, with each specification of a prosody parameter specifying a target value for said prosody parameter, and a point in time for reaching said target value, which point in time is unrestricted to any particular point within said_duration, to thereby generate a signal adapted for converting into speech.

As for claim 18, it depends on claim 1, and it also specifies more than one prosody parameter specification, with each specification having its own target value, and a point in time for reaching this value. Additionally, the time for reaching the target is unrestricted, *a priori* (that is, before the method starts) as to where within the phoneme's duration it is set. Since claim 1, which includes such a specification for at least one prosody parameter, is not taught or suggested by any combination of prior art, it follows that claim 18 is also not taught or suggested by any such combination of the cited prior art.

Claims 19 and 20 depend on claim 18.

Claim 21 is an independent claim. It specifies "a third step for inserting...," which appears to be the subject of the above-mentioned assertion by the Examiner that this third step is taught by Yang et al at col. 5, lines 1-12. Applicants respectfully disagree.

The cited col. 5, lines 1-12 text states

> The synchronization adjusting unit 14 receives the processing results from the prosody processing unit 13, and adjusts the time durations for every phoneme to synchronize the image signal by using the synchronization information which was received from the multi-media distributor 11. With the adjustment of the time duration of phonemes, the lip shape can be allocated to each phoneme in accordance with the position and manner of articulation for each phoneme, and the series of phonemes is divided into small groups corresponding to the number of the lip shapes recorded in the synchronization information by comparing the lip shape allocated to each phoneme with the lip shape in the synchronization information.

Beutnagel 3-12-9

This is a teaching that the time durations of the phonemes that are determined by processing unit 13 are not employed without alteration. Rather, unit 14 modifies those time duration values so as to synchronize every phoneme to image signals. This says nothing about creating any time offsets from beginning of a phoneme, except to the extent that the Examiner equates a time offset with a duration specification. That, however, is NOT what applicant's claims specify.

The third step of claim 21 (in the unamended claim) specifies (a) a target prosody parameter value, (b) within the duration of a phoneme, (c) that value being at an explicitly chosen time (rather than dictated a priori by the algorithm), (d) this time being a time offset from the beginning of the phoneme's duration, and (e) the time being more than zero and less than the duration of the phoneme. None of these attributes are taught or suggested by Yang et al, or by Yang et al in view of Campbell et al. An assertion that is much weaker, but it still dictates the conclusion that claim 21 is patentable is that no combination of the cited references teaches or suggests all of these attributes. Therefore, it is earnestly believed that claim 21, even in its unamended form) is clearly patentable.

Nevertheless, claim 21 is amended herein with the hope that these attributes will come into focus more clearly for the reader. As amended, claim 21 still specifies (a) a prosody parameter target value that the prosody parameter is to reach within (b) a duration of the phone, and the step is enabled to (c & d) explicitly specify a time offset from the beginning of the duration of said phoneme. Additionally, amended claim 21 is even more explicit by adding that (e) the explicitly specified time is greater than and less than the duration of said phoneme.

In short, applicants respectfully submit that amended claim 21 is not obvious over Yang et al in view of Campbell et al.

In light of the fact that various known references either do not specify a time for reaching a target value, or specify a time that is rigidly preassigned by the algorithm to occur at a particular instant, applicants respectfully submit that there is no suggestion, and no motivation, anywhere in various known references to modify their algorithms so that the instant at which a prosody parameter reaches its target value is specified by an explicit data field.

Beutnagel 3-12-9

The advantage that such a capability imparts was not realized by any of the cited references, and since such a capability requires the inclusion of an additional field (and the attendant information-carrying burden) there would have been no reason for these references to take the step taken by applicants.

Claim 22 defines the behavior of a defined prosody parameter, and there is absolutely no teaching regarding this aspect in either Yang et al or in Campbell et al, and the Examiner has not pointed to any. Therefore, in addition to the fact that claim 22 depends on claim 21, applicants believe that claim 22 defines subject matter that is not suggested by Yang et al in combination with Campbell et al.

As for claims 2, 3, 5, 15, 16, applicants note that these claims depend on claim 1, which applicants believe to be patentable over the Yang et al and Campbell et al combination of references. Therefore, claims 2, 3, 15, and 16 are also believed patentable.

As for claims 7 and 17, applicants respectfully disagree with the Examiner that Yang et al teach target values. A target value is more than just value. A target value is a value to which one approaches, and it takes time to approach it. A value that is set at the beginning of a phoneme duration is not considered a "target value" but merely a "value."

Applicants also disagree that the Examiner's assertion that in col. 2, lines 55-65 Yang et al teach target specs in terms of offsets from the boundaries; and even if they did, applicants respectfully disagree that it is material to claim 7 in light of the claim 7 limitations. In applicants' view, actually, the cited text speaks of simply adjusting the duration of phonemes. That has nothing to do with explicitly specifying a particular point in time, within the duration of a phoneme, at which time a specified prosody parameter of the phoneme reaches its target value. Applicants respectfully submit, therefore, that claims 7 and 17 are clearly not obvious in light of the Yang et al and Campbell et al combination of references (separate and apart from their dependence on a parent claim).

Claims 10, 13, and 14 ultimately depend on claim 1.

Beutnagel 3-12-9

In light of the above amendments and remarks, applicants respectfully submit that all of the Examiner's rejections have been overcome. Reconsideration and allowance are, therefore, respectfully solicited.

Respectfully,
Mark Beutnagel
Joern Ostermann
Schuyler Quackenbusch

Dated: 10/9/03

By _____
Henry T. Brendzel
Reg. No. 26,844
Phone (973) 467-2025
Fax (973) 467-6589
email brendzel@comcast.net